Fundamental frequency estimation of low-quality electroglottographic signals

Christian T. HERBST [1], Jacob C. DUNN [2,3]

[1] *Bioacoustics Laboratory, Department of Cognitive Biology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria*

[2] *Department of Animal and Environmental Biology, Faculty of Science & Technology, Anglia Ruskin University, Cambridge, UK*

[3] *Division of Biological Anthropology, University of Cambridge, Cambridge, UK*

Corresponding author: Christian T. HERBST
email: herbst@ccrma.stanford.edu

Keywords: electroglottography; EGG; fundamental frequency; f_o; glottal closure instant; GCI; Praat; SIGMA

Abstract

Fundamental frequency (f_o) is often estimated based on electroglottographic (EGG) signals. Due to the nature of the method, the quality of EGG signals may be impaired by certain features like amplitude or baseline drifts, mains hum or noise. The potential adverse effects of these factors on f_o estimation has to date not been investigated.

Here, the performance of thirteen algorithms for estimating f_o was tested, based on 147 synthesized EGG signals with varying degrees of signal quality deterioration. Algorithm performance was assessed through the standard deviation σ_{fo} of the difference between known and estimated f_o data, expressed in octaves.

With very few exceptions, simulated mains hum, and amplitude and baseline drifts did not influence f_o results, even though some algorithms consistently outperformed others. When increasing either cycle-to-cycle f_o variation or the degree of subharmonics, the SIGMA algorithm had the best performance (max. σ_{fo} = 0.04). That algorithm was however more easily disturbed by typical EGG equipment noise, whereas the NDF and Praat's auto-correlation algorithms performed best in this category (σ_{fo} = 0.01).

These results suggest that the algorithm for f_o estimation of EGG signals needs to be selected specifically for each particular data set. Overall, estimated f_o data should be interpreted with care.

Introduction

Fundamental frequency (f_o) is one of the key parameters used for the quantitative description of voice signals (Baken & Orlikoff 2000; [1]–[4]. f_o represents the rate of vibration of the laryngeal sound generator, typically consisting of the vocal folds in humans and most mammals. f_o detection is performed under the assumption that the analyzed sound source exhibits periodic vibration.

A time series such as the (acoustic) voice signal is said to be periodic when it precisely repeats itself at certain intervals, mathematically expressed as

$$x(t \pm nT_o) = x(t) \tag{1}$$

where t is time, n is a positive integer and T_o is the period [5], i.e., the duration of one glottal cycle. The smallest possible value of T_o of a periodic time series that satisfies Eq. 1 is called the fundamental period of that time series, and its inverse is the fundamental frequency:

$$f_o = \frac{1}{T_o} \tag{2}$$

Several different ways of denoting fundamental frequency are used in the literature (such as, e.g., *F0* or f_o with a subscript zero). However, a recent consensus paper suggests to use the denotation f_o with a lower letter *f* and the character $_o$ (for "oscillatory") instead of the zero (for "zero harmonic") as a subscript [6].

f_o is often confused with "pitch". f_o is a property of the vibration of a physical system, measured in Hertz [Hz]. In contrast, pitch is a psychoacoustic quantity, defined as "*that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to* *high*” [7]. In quite a few cases the two quantities approximate each other, but not always. Hence, the term “pitch” should only be used if (human) perception is addressed, and be avoided when laryngeal sound generation is described as a physical system.

Voice, as practically any other biosignal, is never purely periodic. Rather, it is nearly-periodic at best (some authors use the term “quasi-periodic”, which, however, is reserved for describing a signal with two individual fundamental frequencies [5], [8]). For one, f_o traces typically contain linear or quadratic terms, introduced by gradual changes of f_o. Additionally, even the most steady vocalizations contain slight cycle-to-cycle alterations - see [9] for a very good discussion. More severe phenomena are constituted by irregularity/chaos, subharmonics (“period doubling”, “period tripling”, etc) or multiphonia or biphonation, constituted by two independent sound sources [10], [11]. These issues make f_o detection non-trivial, particularly so pathologic voices, certain singing styles, and in animal bioacoustics, where often the laryngeal sound source exhibits non-linear phenomona like irregularity, subharmonics, and bifurcations between different vibratory states [12].

Strictly speaking, f_o can thus not be calculated for voice signals, because f_o is a property of purely periodic signals. Consequently, there is always a certain degree of inherent inaccuracy in any f_o estimation. In the words of Owren and Linker, “*All pitch extraction techniques are found to fail under some circumstances, which places a burden on the investigator to consistently monitor the performance of each routine being used*” (1995). Regrettably, apart from some informal recommendations [13], no rigorously established limit or respective error ranges for the acceptable degree of irregularity have been established. This makes comparison of f_o data ranges presented in different studies highly problematic.

One additional complication of f_o detection is sometimes introduced by the degeneration of the analyzed acoustic signal by background noise. Lacking anechoic chambers or other adequately sound-treated rooms, in a medical setting this problem can be circumvented by directly assessing the process of laryngeal sound production, e.g., via the glottal area waveform [14], derived by analysis of endoscopic laryngeal high-speed videos [15], [16]. However, the respective equipment is expensive and not always available.

A non-invasive alternative for assessing laryngeal vibration is electroglottography (EGG), pioneered by Fabre in 1957 [17]. In EGG, a high-frequency, low-voltage current is passed between two electrodes, which are placed on either side of the thyroid cartilage. Changes in vocal fold contact area during vocal fold vibration result in admittance variations, and the resulting EGG signal is proportional to the relative vocal fold contact area (VFCA) [18]. A number of parameters quantitatively describing the laryngeal sound generation process can be extracted from a properly recorded EGG signal [19]. Amongst others, the EGG signal is an ideal candidate for assessment of the (time-varying) f_o because it is neither influenced by vocal tract acoustics nor by background noise.

Even under optimal conditions there can be a certain degree of distortion in an acquired EGG signal – see e.g. [20] for a discussion. Further quality degeneration of the EGG signal can be introduced by inadequately positioned EGG electrodes (e.g. caused by excessive larynx or neck movement); reduced conductivity between EGG electrodes and the larynx due to tissue fat, beards, or fur (in animals); radio signals or mains hum interference with the utilized electroglottograph; or noise introduced during (potentially wireless) transmissions of the EGG signals from the electroglottograph to the recording device. Furthermore, EGG signals from pathologic voice production can consist of non-periodic sequences. All these influences potentially pose challenges for f_o detection, as described above.

For these reasons, we decided to formally assess the performance of a number of algorithms for f_o estimation when analyzing a set of synthesized EGG signals with six types of artificially induced distortion of quality. One key aim of this study is to assess how the Praat software package, one of the standard tools for f_o assessment in animal bioacoustics, performs in relation to eight other algorithms mainly known from human speech processing.

Materials and Methods

Synthetic test signals

A set of synthesized EGG signals at various stages of corruption were generated at a sampling frequency of 48000 Hz. Each synthesized signal had a duration of 2 seconds. The f_o information for each glottal cycle within a synthesized signal was derived randomly from a gaussian distribution centered around 1000 Hz with a standard deviation of 500 Hz. Only f_o data between 100 Hz and 2000 Hz were considered. This extended range was chosen to encompass the singing voice range of humans, and vocalizations of some non-human mammals.

The f_o values were sorted in ascending order, and the resulting information was used to drive a kinematic vocal fold vibration model [21]. The model's default parameters were used (Q_a = 0.3; Q_s = 3.0; Q_b = 1.0; Q_p = 0.2). This process resulted in synthetic EGG signals with non-linearly increasing f_o, as illustrated in Figure 1. The time offset and the period of the resulting glottal cycles within each synthesized signal were stored for later comparison with the analysis results from the tested algorithms.

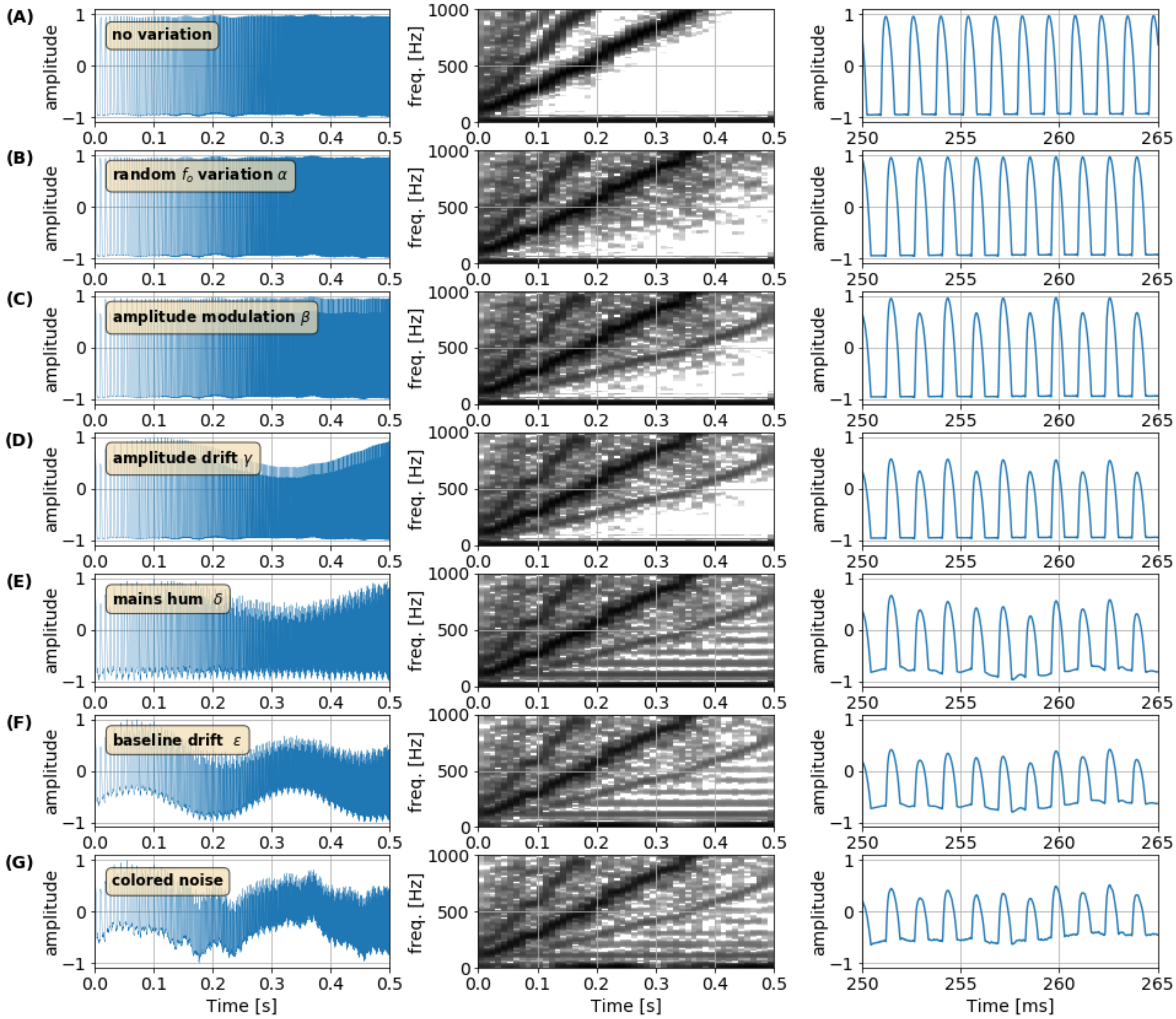

Figure 1: Illustration of EGG signal synthesis, cumulatively introducing the features which degenerate the EGG signal quality at various stages (see text). The left panels show the EGG signal (reduced to 0.5 s to increase the clarity of the illustration). The middle panels contain a narrow-band spectrogram of the EGG signal. In the right panels the first four glottal cycles of each signal are displayed. (A) undistorted synthesized signal; (B) random fo variation, a = 0.15; (C) introduction of subharmonics, b = 0.15; (D) amplitude drift added, g = 0.3; (E) mains hum added, d = 0.15; (F) baseline drift added, e = 0.4; (G) typical EGG equipment noise added, SNR = 15 dB.

As mentioned in the introduction, there are a number of factors that can introduce distortions into the recorded EGG signal and make f_o estimation problematic. In order to test the potential effect of these factors, the following features were introduced into the synthesized EGG signals at various degrees:

1. **Random f_o variation**: When generating the individual EGG cycles, their respective period was allowed to vary randomly within a certain range. This processing step was introduced after sorting the f_o values retrieved from the Gaussian distribution (see above). The final f_o of consecutive cycles within each synthesized signal was determined by

$$f_o^{'}(t) = f_o[1 + \alpha(RND[0..1] - 0.5)] \quad (3)$$

where α is the f_o random factor, which was varied between 0 (no f_o variation) and 0.3. A comparison between α and the relative average perturbation (RAP), a voice quality parameter to assess pathologic human voice production [22], suggests a relationship of $RAP = 0.2118\ \alpha + 0.0029$, $R_2 = 0.9996$ (The y-intercept of 0.0029 was introduced by the non-linear increase of f_o in the synthesized signals). As a reference, for healthy humans RAP values of 0.0021 to 0.0089 were reported [19], which would be the equivalent of α = [-0.0038..0.0283]. Pathologic voices were measured to have RAP values of 0.0068 to 0.0452, corresponding to α = [0.0187..0.1997].

2. **Subharmonics**: The presence of subharmonics, a relatively common feature in mammalian vocalization, was simulated by scaling the amplitude of every other synthesized EGG glottal cycle by (1 - β), where the factor β was varied between 0 and 0.3. Non-zero values of β resulted in the appearance of period-2 subharmonics (period doubling). The parameter value range follows Bergan & Titze [23], who found that the perceptual pitch-drop of an octave occurred at amplitude modulation rates of 10 – 30 %.

3. **Amplitude drift**: The temporal variation of the EGG signal amplitude was simulated by introducing a sinusoidally varying amplitude modulation at an arbitrarily fixed rate of f_{AM}= 2.27 Hz. In particular, the EGG signal was multiplied by

$$(1-\gamma)+\frac{\gamma[1+\sin(2\pi f_{AM}\,t)]}{2} \qquad (4)$$

where the amplitude modulation factor γ was varied between 0 and 0.6 between across synthesized signals.

4. **Mains hum**: A mains hum signal with a duration of 2 s was synthesized as a harmonic series at 100 Hz f_o with a steadily decaying spectral envelope, using a spectral slope of -6 dB per octave. A total of 20 harmonics were included. This mains hum signal was scaled by the factor δ and then added to the normalized synthesized EGG. The amplitude scaling factor δ was varied between 0 (no mains hum added) and 0.3.

5. **Baseline drift**: The baseline offset of the signal was allowed to vary sinusoidally at an arbitrarily defined fixed rate. In particular, the following baseline drift was added to the synthesized signals: $0.5\ \varepsilon\ [1+sin(2\pi f_{BD}t)]$, where f_{BD} = 1.71 Hz. The factor ε was varied between 0 (no baseline drift) and 0.8 across the synthesized signals.

6. **Noise**: Finally, colored noise was added to the synthesized EGG signal in order to simulate various signal-to-noise ratios (SNR). SNR was varied in a range of -5 dB to 35 dB. These values were taken from a recent study which reported these surprisingly low SNR ranges for EGG signals recorded from humans in laboratory conditions [24]. The noise was generated by scaling the frequency components of white noise in the frequency domain during a forward-backward Fourier transform. The amplitudes for the frequency-dependent scaling were derived from averaged noise contained in previously recorded EGG signals [24].

Deviating from the "best case" of the synthesized EGG signal, six data sets were generated, where each of the aforementioned six parameters was varied in isolated fashion at 21 equidistantly spaced intervals – see Figure 1 for an example. Additionally, one data set was generated where all six parameters were varied at once (termed "compound" scenario in the remainder of this text). In this fashion, a total of 147 (21 x 7) synthesized EGG signals were produced.

Evaluated algorithms

Owing to frequently occurring linear/quadratic trends and cycle-to-cycle aberrations in a voice signal, quantitative analysis focuses on the time-varying rather than the mean f_o. This is either achieved by short-term windowed approaches [25], where short portions of the voice signal are evaluated at consecutive time instants, or by estimation of the so-called glottal closure instants (GCI) [26]. GCI operates on the assumption that the major sound generation event occurs at the instant of glottal closure, i.e., (partial) collision of the laryngeal or syringeal tissue, and that each glottal cycle has a period (sometimes called "epoch" - [27]) that is determined by two consecutive GCIs. Recalling Eq. 2, the time-varying f_o is then found by taking the inverse of the period.

f_o estimators and GCI detectors may operate on different computational principles. The majority of them are rooted in either the time domain (looking at similarities or recurrent features in the voice signal) or in the frequency domain (by further analyzing the time-varying spectrum of the voice signal, as produced by a Fourier transform). Alternative approaches include, e.g., wavelet analysis [28]–[30] or phase space analysis [31]. Description of the concepts involved in the various algorithms is beyond the scope of this manuscript. The reader is referred to the landmark textbook by Hess [25]. Some key approaches are described in Owren & Linker's summary [4]. Further good overviews are given by Talkin [32] and Drugman *et al*. [33], the latter of which discusses some more recent developments.

Overall, a surprisingly large number of f_o estimators have been described in the literature. Already in 1983 Hess states that "*literally hundreds of pitch-determination methods and algorithms have been developed*" [25]. In the supplementary materials of this manuscript we include a non-exhaustive list of 75 f_o and GCI estimators, addressing some past and recent developments, and providing web links to free source code or software applications where applicable.

Given the number of options, a practical solution had to be found for selecting the algorithms tested in this manuscript. Besides focusing on the algorithms included in the Praat software package [34], our main selection criterion was (a) free availability of the algorithm source code, and (b) ability to operate the algorithm within a free software environment, i.e., the Linux operating system and, where applicable, GNU Octave [35], the free equivalent to Matlab. A total of thirteen such algorithms were included in this study:

- five algorithms from the Praat software, version 5.4.06. The following methods were tested in this study: “to Pitch (ac)”, “to Pitch (SHS)”, “to PointProcess (periodic, cc)”, “'to PointProcess (periodic, peaks)”, and “to PointProcess (zeroes)”. These algorithms are referred to as **Praat (AC)**, **Praat (SHS)**, **Praat (periodic cc)**, **Praat (periodic peaks)**, and **Praat (zeros)**, respectively, for the remainder of this text. Preliminary analysis suggested that Praat's methods “to Pitch (SPINET)” and “to PointProcess (extrema)” produced greatly inferior results. These two algorithms were thus excluded from this report;
- the **DECOM** algorithm [36] presented in [37];
- the **DYPSA** GCI algorithm, introduced by Kounoudes *et al*. [38] and described in further detail by Naylor *et al*. [39];
- The **NDF** (Nearly Defect-Free) f_o detector [40], implemented as `MulticueF0v14.m`, version 2016-06-30;
- The **RAPT** algorithm by [32], implemented as `fxrapt.m` in the voicebox package;
- David Talkin's **REAPER** algorithm (unpublished work – see https://github.com/google/REAPER);
- the **SIGMA** GCI detector, developed by Thomas & Naylor [30];
- the **SWIPE'** algorithm, developed by Camacho & Harris [41];
- the **YAGA** GCI detector, developed by Thomas *et al.* [42];

Web links for downloading the software of the algorithms utilized in this comparison can be found in the supplementary materials.

All algorithms were controlled through a set of custom scripts written in Python by author CTH, operated on Linux 16.04 LTS. Praat and the compiled C-code of REAPER were accessed through command-line pipes. All other algorithms were available as Matlab code. They were thoroughly tested in GNU Octave 4.0 and were then embedded into the custom Python code through Python's oct2py wrapper module for Matlab/Octave code [43]. For all algorithms, the respective standard parameters were used, except for the upper and lower limits, which were (where possible) specified as 100 Hz and 2000 Hz, respectively. The upper frequency limits of REAPER and the voicebox-based DYPSA, RAPT, and SIGMA algorithms had to be changed from 500 Hz to 2000 Hz in the respective source code. All f_o detection algorithms (Praat (AC), Praat (SHS), NDF, RAPT, REAPER, SWIPE') were operated at a time step of 1 ms.

Combining algorithm outputs

Preliminary assessment of the performance of the algorithms suggested that there was no single algorithm that performed best under all conditions. Rather, the SIGMA GCI detector and the Praat autocorrelation (AC) f_o estimator showed the most robust performance in different subsets of the synthesized data (see Results). In an attempt to consolidate the benefits of these two algorithms, a custom analysis approach (denoted as CUSTOM for the remainder of this manuscript) was implemented as follows: SIGMA GCI data was converted to f_o information at a time-step of 1 ms. For each data point (totalling 2000 for two seconds of synthesized sound), the difference between f_o data from Praat AC and SIGMA was computed, expressed in octaves. If that difference was below a certain threshold, an f_o data point was generated by the CUSTOM algorithm (NaN otherwise). The threshold was arbitrarily defined as 5 % of an octave. Preliminary tests with a more rigoros threshold of 1/120 octave (i.e., 10 musical cents), which approximates the just-noticeable difference for pitch perception in humans [44], considerably decreased the usefulness of the CUSTOM algorithm, due to the great number of rejected data points even at slight levels of EGG signal quality degeneration.

Testing procedure

Including the CUSTOM algorithm, 14 algorithms were tested on the 147 EGG signals described above, resulting in a total number of 2058 observations. Prior to f_o calculation and GCI detection, the EGG signals were band-pass filtered twice using a 3rd order Butterworth filter with cutoff frequencies at 20 Hz and 4800 Hz. The 2nd consecutive application of the filter was performed on the time-inverted input signal, in order to negate phase distortion effects. The application of the band-pass filter was deemed appropriate, because comparable pre-processing steps would be performed in "real" data analysis situations. The cutoff frequencies were chosen carefully so as not to distort the analyzed signals.

Evaluation of performance

The output of f_o estimators and GCI detectors is fundamentally different in nature. While the f_o estimators produce equidistantly spaced data points (every 1 ms in the case of this study) representing the time-varying (quasi-instantaneous) f_o information, the GCI detectors provide estimates of the time offsets of presumed glottal closure instants. In order not to add any bias to the analysis (neither in favor of either f_o estimation nor GCI detection methods), we initially decided to compare the performance of all tested algorithms in both domains.

When comparing two frequencies, their difference in Hertz is meaningless as an absolute value. A relative measure needs to be established instead. For the purpose of this study, the frequency differences between known and estimated f_o values were expressed in octaves [45]:

$$\Delta oct = \log_2\left(\frac{f_{SYNTH}}{f_{EST}}\right) \qquad (5)$$

For performance evaluation in the f_o domain, the glottal cycle information from the systhesized signal was converted to a time-series of f_o data at intervals of 1 ms. Based on this information, the following three parameters were calculated:

- A success metric, expressing the number of produced f_o data points in percent:

$$\rho_{f_o} = 100\frac{m}{n} \tag{6}$$

where n is the total number of possible data points (2000 for two seconds of synthesized sound) and m is the number of actually detected data points.

- Applying Eq. 5, the average of the absolute differences between known f_o information from the synthesized signals and estimated f_o data was computed as follows:

$$\mu_{f_o} = \frac{1}{n}\sum_{0}^{n-1}\left|\Delta oct[i]\right| \tag{7}$$

- Similarly, the standard deviation of f_o estimation was computed as:

$$\sigma_{f_o} = \sqrt{\frac{1}{n}\sum_{0}^{n-1}\left(\Delta oct[i]\right)^2} \tag{8}$$

The performance metrics parameters ρ_{GCI}, μ_{GCI}, and σ_{GCI} for GCI detection were calculated in analogy to those for f_o estimation, with the difference that n was defined as the total number of glottal cycles in the respective synthesized signal.

Preliminary inspection of the algorithm performance data revealed no remarkable differences between the f_o-based (ρ_{fo}, μ_{fo}, and σ_{fo}) and the respective GCI-based values (ρ_{GCI}, μ_{GCI}, and σ_{GCI}), suggesting that conversion between f_o and GCI information did not introduce noteworthy artifacts into the data. Furthermore, there were no substantial differences of trends between the μ_{fo} and σ_{fo} parameters. For these reasons, the remainder of this text focuses on the f_o-related parameters ρ_{fo} and σ_{fo} alone.

Results

Detailed results of f_o detection from one representative signal are shown in Figure 2. An overview of the parameters ρ_{fo} and σ_{fo} for all analyzed scenarios is given in Figures 3 and 4. Detailed μ_{fo} success rates and σ_{fo} scores for all analysis scenarios are provided in supplementary tables 1 and 2.

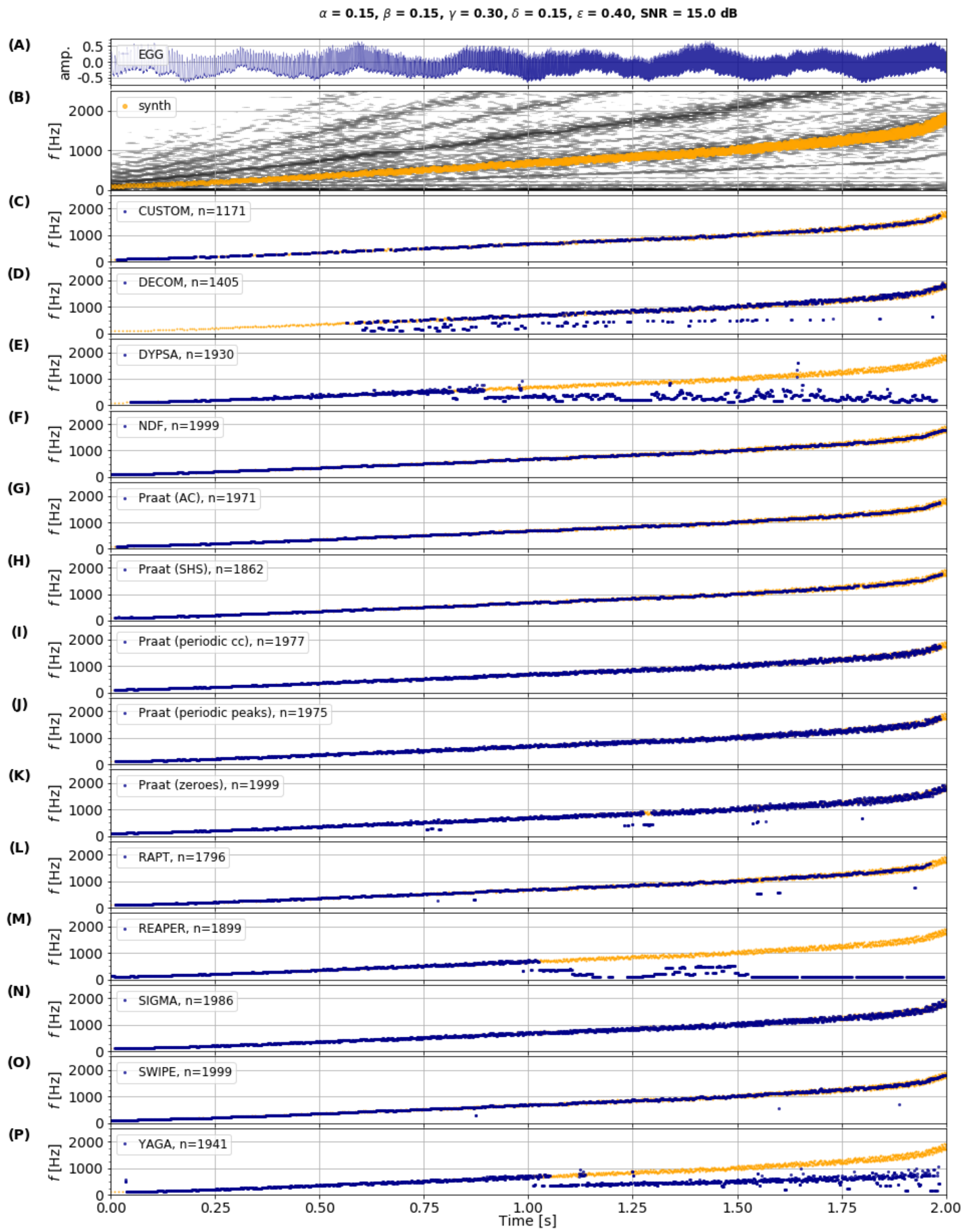

Figure 2: ***Detailed results of f_o detection from the synthesized signal depicted in Figure 1.*** *(A) synthesized EGG signal; (B) narrow-band spectrogram of synthesized EGG signal, known f_o data superimposed. (C) – (P) f_o detection results for all evaluated algorithms (dark dots), superimposed upon known f_o data (light dots). Data from GCI detectors were converted to equidistantly spaced f_o values (see Methods). The illustrated synthesized EGG signal represents the "compound" case 10 in Figures 3G and 4G.*

365

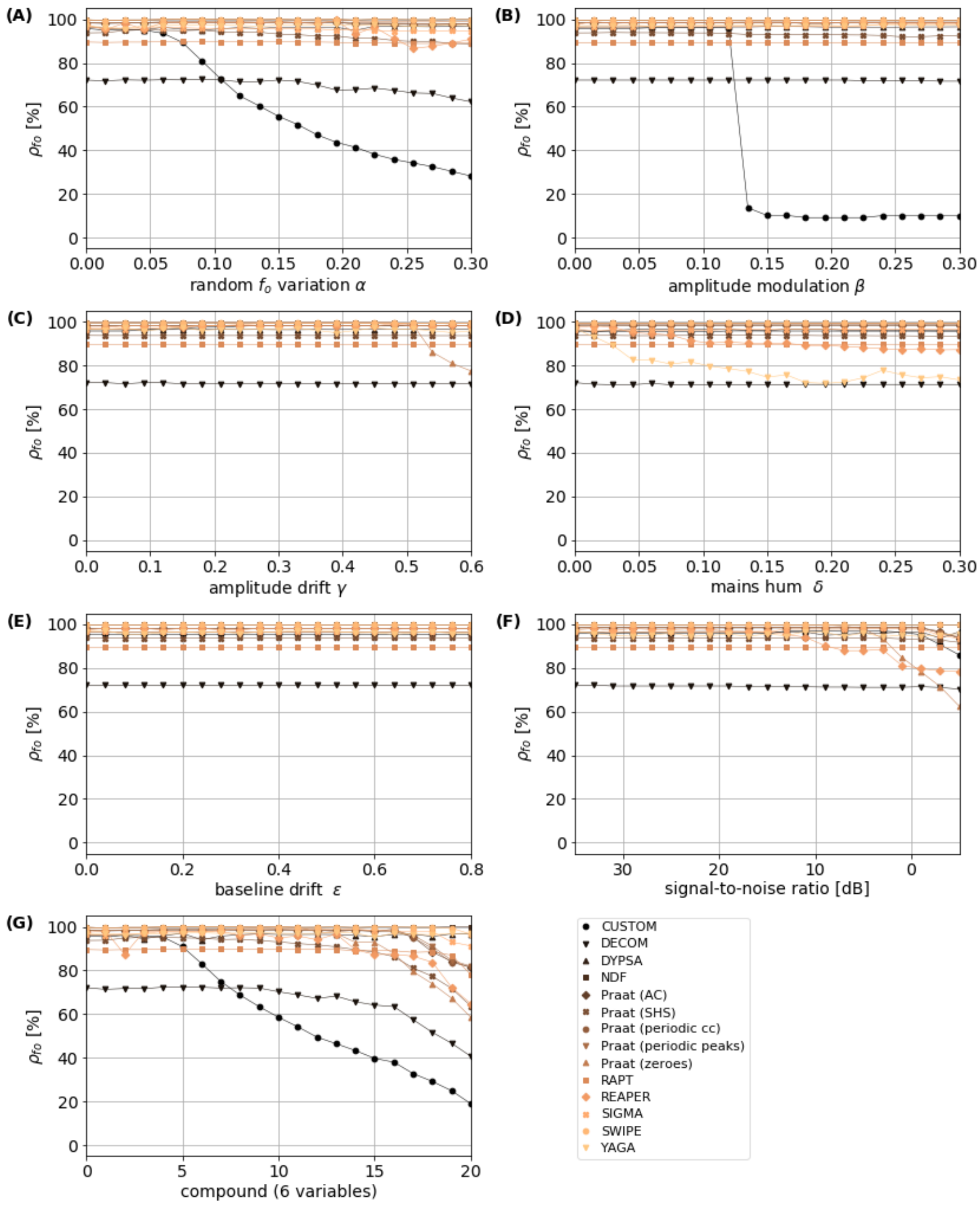

Figure 3: ***f_o data point resolution metric ρ_{fo} for all analyzed algorithms and all synthesized EGG signals****. (A) – (F) ρ_{fo} as a function of the six simulated influence factors on EGG signal quality. (G) effect of simultaneous change of all six influence factors.*

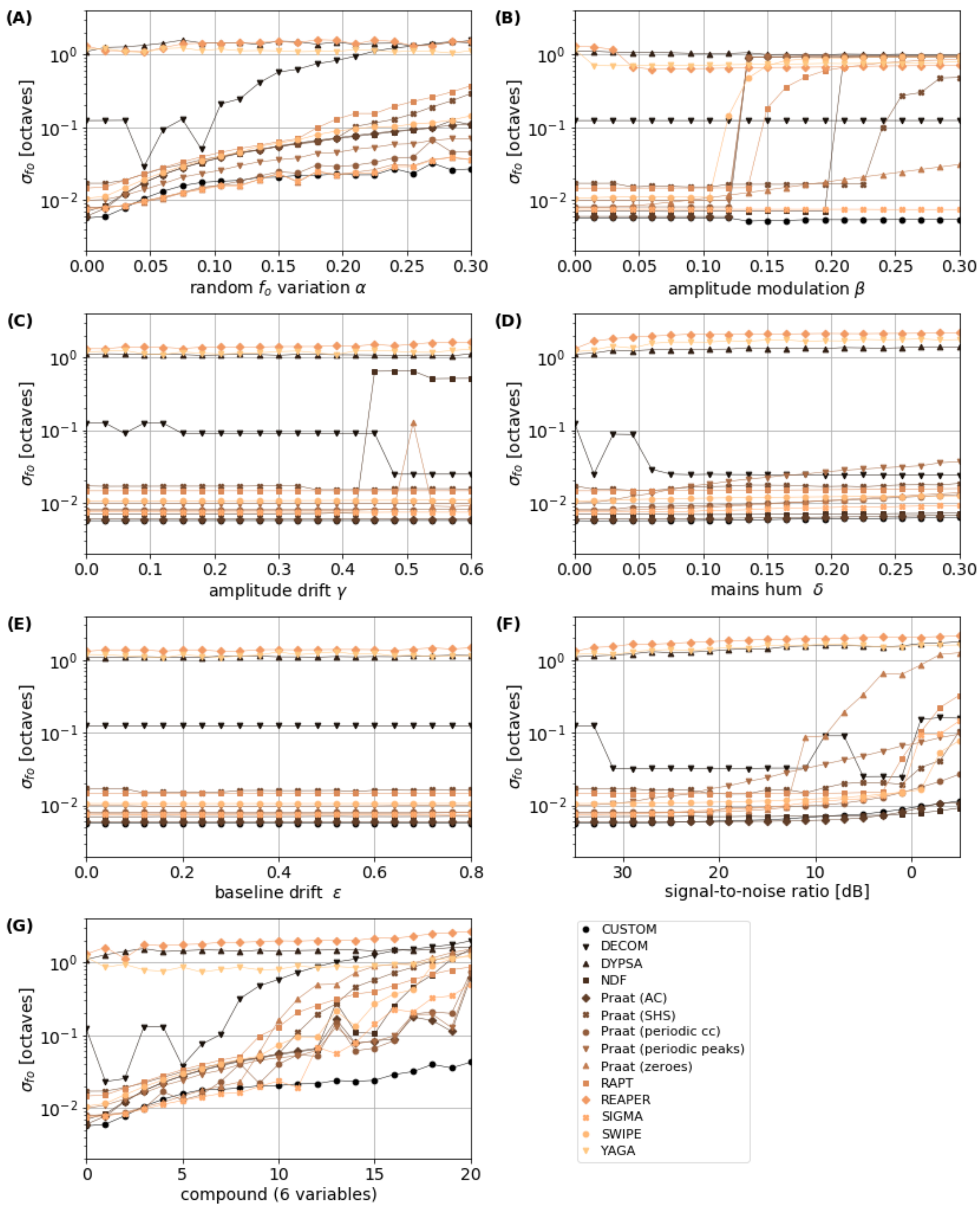

Figure 4: ***f_o detection performance metric σ_{fo} for all analyzed algorithms and all synthesized EGG signals****. (A) – (F) σ_{fo} as a function of the six simulated influence factors on EGG signal quality. (G) effect of simultaneous change of all six influence factors.*

With a few exceptions of extreme EGG signal modifications in the "compound" scenario and for extreme SNR values, most algorithms produced data for more than 90 % of the possible 2000 data points per synthesized signal (see Figure 3). Exceptions to this trend were found in the RAPT and DECOM algorithms, which typically had ρ_{fo} values of about 90 % and 80 %, respectively. The CUSTOM algorithm deviated from its typical 95 % f_o detection success rate when the random f_o variation α was increased above 0.1 and when the amplitude modulation β was greater than 0.12, suggesting that above these critical values the f_o readings from the two algorithms upon which the CUSTOM algorithm is based (i.e., Praat's "to Pitch (AC)" and SIGMA – see Methods) deviated by more than 5 % of an octave.

Three of the analyzed algorithms (DYPSA, REAPER, and YAGA), all designed with the purpose of analyzing human speech, had problems recognizing f_o above ca. 1000 Hz. Consequently, they were the worst performing algorithms analyzed. The error benchmark σ_{fo} for the DECOM algorithm was typically around 10 % of an octave, rising considerably with increased random f_o variation α. All other algorithms started out with acceptable σ_{fo} ratings for EGG signals at lesser degrees of EGG signal quality distortion. However, increased random f_o variation had a tendency to gradually increase σ_{fo} in all algorithms except DYPSA, REAPER, and YAGA. Overall, the CUSTOM and SIGMA algorithms had the best performance when testing for random f_o variation – see Figure 4A.

For most of the algorithms, the occurrence of subharmonics appeared to be a crucial factor which led to abrupt increases in σ_{fo} over an amplitude modulation range of $0.1 > \beta > 0.24$ (see Figure 4B). In each of these cases, the respective algorithm started to latch on to the subharmonic energy components in the signal. The respective threshold values were found at: NDF: $\beta = 0.21;$ Praat (AC): $\beta = 0.14;$ Praat (SHS): $\beta = 0.24;$ Praat (periodic cc): $\beta = 0.14;$ Praat (periodic peaks): $\beta = 0.14;$ RAPT: $\beta = 0.15;$ and SWIPE: $\beta = 0.12.$ As with random f_o variation, the CUSTOM and SIGMA algorithms had the best performance with increased amplitude modulation.

No noteworthy trends were found with variation in amplitude drift, mains hum, or baseline drift – see Figures 4C-E (Preliminary experiments conducted without bandpass filtering the signals before analysis revealed the same trends, even for baseline drifts). The only exception was the NDF algorithm, which suffered an abrupt decrease in performance for amplitude drifts $\gamma < 0.42$, and the DECOM algorithm, which achieved reduced σ_{fo} values for $\gamma < 0.45$ and $\delta < 0.04$.

Finally, typical EGG equipment noise seemed to be an important factor, influencing a number of algorithms – see Figure 4F: There was an almost linear correlation between SNR of noise and σ_{fo} in the "Praat (periodic peaks)" algorithm. More abrupt degenerations of performance (measured by increasing σ_{fo}) were found for the following algorithms at respective thresholds: Praat (SHS): SNR = -5 dB; Praat (zeroes): SNR = 11 dB; RAPT: SNR = 1 dB; SIGMA: SNR = -1 dB; SWIPE: SNR = -3dB. The CUSTOM, NDF, and Praat (AC) algorithms appeared to perform particularly well under the influence of noise, with terminal values of σ_{fo} = 0.01 at an SNR of -5 dB.

Algorithm performance for linear combinations of the six influence factors described above are shown in the "compound" scenario illustrated in Figure 4G. The CUSTOM algorithm had a notably better performance (i.e., lower σ_{fo} values) than all other algorithms, particularly at higher degrees of EGG signal deterioration. This performance success was, however, counterbalanced by the algorithm's lowered success rates μ_{fo} – see Figure 3G.

Discussion

This study examines the performance of a number of f_o and GCI detection algorithms when analyzing a special class of signals, i.e., EGG signals with increasing complexity and at various stages of signal quality degradation. A total of six influence factors were assessed in this study: two inherent to the voice signal itself (random f_o variation and subharmonics), and four types of signal degradations (amplitude and baseline drifts, mains hum, and typical EGG equipment noise). Mains hum, amplitude drift, and baseline drift all appeared to have a lesser influence on algorithm performance. The opposite was true for the other three factors – alteration in cycle-to-cycle variation (introduced by random f_o variation), subharmonics (introduced by amplitude modulation of odd cycles), and typical EGG equipment noise all had a clear impact on algorithm performance.

The somewhat disquieting main finding of this study is that there does not seem to exist one single "best" algorithm for analyzing EGG signals at various stages of complexity and degradation. For high-quality, low-noise EGG signals (e.g., those typically acquired in excised larynx settings) the SIGMA algorithm seems to be the best choice. In signals with low signal-to-noise ratios (SNR), such as those collected *in vivo* from humans with a certain degree of fat tissue or phonating with incomplete glottal closure [24], or signals with suboptimal EGG electrode placement, the SIGMA algorithm does not appear to be the best choice. In those cases, NDF or Praat's auto-correlation (AC) algorithm would appear to be better suited. However, the performance of both NDF and Praat's AC algorithm is negatively affected by the occurrence of subharmonics.

In an attempt to consolidate these trends, a CUSTOM approach was introduced in this manuscript, combining the virtues of both SIGMA and Praat's AC algorithm. This CUSTOM algorithm showed the best performance overall (particularly in the "compound" scenario), but the improved performance came at the expense of discarding a large proportion of the analyzed data in situations where the outputs of the SIGMA and AC algorithms did not converge. This obvious tradeoff between data quality and quantity can to a certain degree be controlled via the CUSTOM algorithm's threshold setting (see Methods).

Some of the analyzed algorithms are intended to operate on certain types of signals [36]. This may partially explain why the DYPSA, REAPER and YAGA algorithms failed to produce meaningful data outside the typical f_o ranges of human speech. Therefore, inferior performance of an algorithm in this study does not constitute a reason to conclude that the respective algorithm is inferior *per se*.

When studying the literature, the following pattern emerged: Most of the proposed f_o and GCI detection algorithms were introduced by comparing their results with those from some other algorithms (differing across the various studies, and typically basing the tests on different input signals across different studies). Interestingly, in all of these cases the respectively proposed algorithm had comparable or better performance than all other algorithms. Four non mutually-exclusive conclusions can be drawn from this phenomenon:

(a) Owing to progress in the field of engineering the newly introduced algorithms become increasingly better over the years;
(b) Some algorithms work better for a certain type of data (e.g., noisy data [46] or special voice production types [47], [48]) than others;
(c) Different methods of estimating algorithm performance result in different outcomes [49];
(d) The authors of studies might have had a certain (unconscious) *a priori* bias towards their "own" algorithm, which may have influenced them in choosing test data and competing algorithms for their performance tests; or finally,
(e) The authors may have made the mistake to train their algorithm on the chosen test data, leading to an over-specialized algorithm performance which can not be generalized to other data sets.

Surprisingly, even studies which are only concerned with comparing algorithm performance (without introducing a new algorithm) do not converge to identical recommendations [48]–[54], suggesting that estimating algorithm performance might be as complex a task as f_o or GCI detection itself. One way to address this issue is by consensually establishing databases of test signals with known properties. Advancing this notion, we have made all synthesized EGG signals utilized in this study available as supplementary materials.

Some of the considerations concerning standardizing algorithm performance evaluation also apply to this study. The f_o range for synthesized signals was somewhat arbitrarily chosen to be in the range of 100 to 2000 Hz, in consideration of the human singing voice and the vocalizations of some non-human mammals. Furthermore, while three of the parameters for determining the synthesized EGG signals were chosen in relation to known values ranges (random cycle-to-cycle variation α, amplitude modulation β, and SNR), the value ranges of the other three parameters had to be defined in an arbitrary fashion, based on the first author's long-term experience with EGG signals. Different values may naturally lead to different performance evaluation results. This is particularly true for the "compound" case, where all six parameters were varied in unison. In fact, preliminary tests with different, more extreme value ranges produced slightly different trends. For this reason, we have refrained from computing an overall metric of success across all synthesized signals. Such a metric would only apply to the given test data set and could not be generalized.

Conclusion

This study corroborates the insight that f_o detection is highly non-trivial [4], [25], [32]. No single "best algorithm" was found for the special class of signals analyzed in this study. Thus, no recommendation for one single all-purpose f_o detection algorithm can be given. Rather, the nature of EGG data needs to be studied carefully before choosing an appropriate algorithm, and the insights from this study can help with that choice. Such an informed approach is recommended, rather than defaulting to a commonly used algorithm.

Summarizing, some main insights from this study are thus that the researcher should never blindly trust a chosen f_o detection algorithm. *Ex post facto*, computed f_o data should always be assessed "by eye", e.g., via f_o traces superimposed upon narrow-band spectrograms. Furthermore, f_o data reported in the literature should not be taken at face value, particularly if the authors did not disclose (a) which f_o detection algorithm was chosen; (b) how the utilized f_o detection algorithm was chosen; and/or (c) whether (and how) the computed data was double-checked manually. There is an inherent degree of uncertainty and error in such data, due to the difficulties in automated f_o detection described in this manuscript.

Acknowledgements

This work is supported by an APART grant from the Austrian Academy of Sciences (to C.T.H.)

Competing interests

No competing interests declared.

References

[1] J. Fischer, R. Noser, and K. Hammerschmidt, "Bioacoustic Field Research: A Primer to Acoustic Analyses and Playback Experiments With Primates," *Am. J. Primatol.*, vol. 75, no. 7, pp. 643–663, Jul. 2013.

[2] W. T. Fitch and M. D. Hauser, "Vocal production in nunhuman primates: acoustics, physiology, and functional constraints on 'honest' advertisement," *Am. J. Primatol.*, vol. 37, pp. 191–219, 1995.

[3] N. H. Fletcher, "Acoustic Systems in Biology: From Insects to Elephants," *Acoust. Aust.*, vol. 3, pp. 83–88, 2005.

[4] M. J. Owren and C. D. Linker, "Some analysis methods that may be useful to acoustic primatologists," in *Current Topics in Primate Vocal Communication*, E. Zimmermann, J. D. Newman, and U. Jürgens, Eds. New York: Springer, 1995, p. 286.

[5] I. R. Titze, *Workshop on acoustic voice analysis. Summary statement*. National Center for Voice and Speech, 1995.

[6] I. R. Titze, "Some Consensus has been Reached on the Labeling of Harmonics, Formants, and Resonances," *J. Voice*, vol. 30, no. 2, p. 129, Mar. 2016.

[7] ANSI, "USA Standard Acoustical Terminology (Including Mechanical Shock and Vibration)," *Tech. Rep.*, vol. S1.1-1960, 1960.

[8] P. Bergé, Y. Pomeau, and C. Vidal, *Order within chaos: towards a deterministic approach to turbulence*. Paris: Hermann and John Wiley & Sons, 1984.

[9] R. M. Roark, "Frequency and voice: perspectives in the time domain," *J Voice*, vol. 20, no. 3, pp. 325–354, 2006.

[10] H. Herzel, D. Berry, I. R. Titze, and M. Saleh, "Analysis of vocal disorders with methods from nonlinear dynamics," *J Speech Hear Res*, vol. 37, no. 5, pp. 1008–1019, 1994.

[11] I. R. Titze, R. J. Baken, and H. Herzel, "Evidence of chaos in vocal fold vibration," in *Vocal fold physiology: frontiers in basic science*, I. R. Titze, Ed. San Diego, CA: Singular Publishing Group, 1993, pp. 143–188.

[12] W. T. Fitch, J. Neubauer, and H. Herzel, “Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production,” *Anim. Behav.*, vol. 63, pp. 407–418, 2002.

[13] G. Friedrich and P. H. Dejonckere, “Das Stimmdiagnostik‒Protokoll der European Laryngological Society (ELS) - erste Erfahrungen im Rahmen einer Multizenterstudie,” *Laryngo-, Rhino-, Otol.*, vol. 84, no. 10, pp. 744–752, 2005.

[14] D. G. Childers, J. M. Naik, J. N. Larar, A. K. Krishnamurthy, and G. P. Moore, “Electroglottography, speech, and ultra-high speed cinematography,” in *Vocal Fold Physiology and Biophysics of Voice*, I. R. Titze and R. Scherer, Eds. Denver: Denver Center of Performing Arts, 1983, pp. 202–220.

[15] D. D. Deliyski and R. E. Hillman, “State of the art laryngeal imaging: research and clinical implications,” *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 18, pp. 147–152, 2010.

[16] S. Hertegard, “What have we learned about laryngeal physiology from high-speed digital videoendoscopy? ,” *Curr Opin Otolaryngol Head Neck Surg*, vol. 13, pp. 152–156, 2005.

[17] P. Fabre, “Un procédé électrique percuntané d’inscription de l’accolement glottique au cours de la phonation: glottographie de haute fréquence; premiers résultats (A non-invasive electric method for measuring glottal closure during phonation: high frequency glottogr,” *Bull. Acad. Nat. Med.*, vol. 141, pp. 66–69, 1957.

[18] V. Hampala, M. Garcia, J. G. Svec, R. C. Scherer, and C. T. Herbst, “Relationship between the Electroglottographic Signal and Vocal Fold Contact Area,” *J. Voice*, vol. 30, no. 2, pp. 161–171, 2016.

[19] R. J. Baken and R. F. Orlikoff, *Clinical Measurement of Speech and Voice (2nd Edition)*, vol. 2nd Editio. Singular Publishing, Thompson Learning, 2000.

[20] R. J. Baken, “Electroglottography,” *J. Voice*, vol. 6, no. 2, pp. 98–110, 1992.

[21] I. Titze, “A four-parameter model of the glottis and vocal fold contact area,” *Speech Commun.*, vol. 8, no. 3, pp. 191–201, 1989.

[22] Y. Koike, “Application of some acoustic measures for the evaluation of laryngeal dysfunction,” *Stud. Phonol.*, vol. VII, pp. 17–23, 1973.

[23] C. Bergan and I. R. Titze, “Perception of pitch and roughness in voice signals with subharmonics,” *J. Voice*, vol. 15, no. 2, pp. 165–175, 2001.

[24] C. T. Herbst, H. K. Schutte, D. L. Bowling, and J. G. Svec, “Comparing chalk with cheese - The EGG contact quotient is only a limited surrogate of the closed quotient,” *J Voice*, vol. 31, no. 4, pp. 401–409, 2017.

[25] W. Hess, *Pitch Determination of Speech Signals: Algorithms and Devices*. Heidelberg, Germany: Springer-Verlag, 1983.

[26] V. N. Tuan and C. D’Alessandro, “Robust Glottal Closure Detection Using The Wavelet Transform.,” *Proc. Eur. Conf. SPEECH Technol.*, pp. 2805--2808, 1999.

[27] T. Ananthapadmanabha and B. Yegnanarayana, “Epoch extraction from linear prediction residual for identification of closed glottis interval,” *IEEE Trans. Acoust.*, vol. 27, no. 4, pp. 309–319, Aug. 1979.

[28] S. Kadambe and G. F. Boudreaux-Bartels, “Application of the wavelet transform for pitch detection of speech signals,” *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 917–924, Mar. 1992.

[29] C. Manfredi, M. D'Aniello, P. Bruscaglioni, and A. Ismaelli, "A comparative analysis of fundamental frequency estimation methods with application to pathological voices.," *Med. Eng. Phys.*, vol. 22, no. 2, pp. 135–47, Mar. 2000.

[30] M. R. P. Thomas and P. A. Naylor, "The SIGMA Algorithm: A Glottal Activity Detector for Electroglottographic Signals," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 17, no. 8, pp. 1557–1566, Nov. 2009.

[31] M. Hagmüller and G. Kubin, "Poincaré Sections for Pitch Mark Determination," in *ITRW on Non-Linear Speech Processing (NOLISP 05)*, 2005, pp. 1–7.

[32] D. Talkin, "A Robust Algorithm for Pitch Tracking (RAPT)," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. New York: Elsevier, 1995, pp. 495–518.

[33] T. Drugman, P. Alku, A. Alwan, and B. Yegnanarayana, "Glottal source processing: From analysis to applications," *Comput. Speech Lang.*, vol. 28, no. 5, pp. 1117–1138, 2014.

[34] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." Institute of Phonetic Sciences, University of Amsterdam, Amsterdam, The Netherlands, 2017.

[35] J. W. Eaton, D. Bateman, S. Hauberg, and R. Wehbring, "GNU Octave version 4.0.0 manual: a high-level interactive language for numerical computations." 2015.

[36] N. Henrich, "DECOM." 2006.

[37] N. Henrich, C. d'Alessandro, B. Doval, and M. Castellengo, "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation," *J. Acoust. Soc. Am.*, vol. 115, no. 3, pp. 1321–1332, 2004.

[38] A. Kounoudes, P. A. Naylor, and M. Brookes, "The DYPSA algorithm for estimation of glottal closure instants in voiced speech," in *ICASSP*, 2002, pp. 349–352.

[39] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of Glottal Closure Instants in Voiced Speech Using the DYPSA Algorithm," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 1, pp. 34–43, Jan. 2007.

[40] H. Kawahara, A. De Cheveigné, H. Banno, T. Takahashi, and T. Irino, "Nearly Defect-free F0 Trajectory Extraction for Expressive Speech Modifications based on STRAIGHT," in *Interspeech*, 2005, pp. 537–540.

[41] A. Camacho and J. G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 124, no. 3, pp. 1638–1652, Sep. 2008.

[42] M. R. P. Thomas, J. Gudnason, and P. A. Naylor, "Estimation of Glottal Closing and Opening Instants in Voiced Speech Using the YAGA Algorithm," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 20, no. 1, pp. 82–91, Jan. 2012.

[43] E. Jones, T. Oliphant, P. Peterson, and Others, "SciPy: Open Source Scientific Tools for Python." .

[44] T. Rossing, *The Science of Sound*. Addison-Wesley Publishing Company, 1990.

[45] R. W. Young, "Terminology for Logarithmic Frequency Units," *J. Acoust. Soc. Am.*, vol. 11, pp. 134–139, 1939.

[46] T. Drugman, T. Drugman, and A. Alwan, "Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics," in *Interspeech*, 2011.

[47] J. Kane and C. Gobl, “Evaluation of glottal closure instant detection in a range of voice qualities,” *Speech Commun.*, vol. 55, no. 2, pp. 295–314, 2013.

[48] O. Babacan, T. Drugman, N. Henrich, and T. Dutoit, “A Quantitative Comparison of Glottal Closure Instant Estimation Algorithms on a Large Variety of Singing Sounds.” .

[49] L. Rabiner, M. Cheng, A. Rosenberg, and C. McGonegal, “A comparative performance study of several pitch detection algorithms,” *IEEE Trans. Acoust.*, vol. 24, no. 5, pp. 399–418, Oct. 1976.

[50] S.-J. Jang, S.-H. Choi, H.-M. Kim, H.-S. Choi, and Y.-R. Yoon, “Evaluation of Performance of Several Established Pitch Detection Algorithms in Pathological Voices,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, vol. 2007, pp. 620–623.

[51] M. J. Cheng, L. R. Rabiner, A. E. Rosenberg, and C. A. McGonegal, “Comparative performance study of several pitch detection algorithms,” *J. Acoust. Soc. Am.*, vol. 58, no. S1, pp. S61–S62, Nov. 1975.

[52] V. Parsa and D. G. Jamieson, “A comparison of high precision F0 extraction algorithms for sustained vowels.,” *J. Speech. Lang. Hear. Res.*, vol. 42, no. 1, pp. 112–26, Feb. 1999.

[53] I. R. Titze and H. Liang, “Comparison of F0 extraction methods for high precision voice perturbation measurements,” *J. Speech Hear. Res.*, vol. 36, no. 6, Dec, pp. 1120–1133, 1993.

[54] A. Tsanas, M. Zañartu, M. A. Little, C. Fox, L. O. Ramig, and G. D. Clifford, “Robust fundamental frequency estimation in sustained vowels: Detailed algorithmic comparisons and information fusion with adaptive Kalman filtering,” *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2885–2901, May 2014.

Supplementary materials to: Herbst & Dunn, Fundamental frequency estimation of low-quality electroglottographic signals

November 23, 2017

Overview of published algorithms for fundamental frequency detection

The following table provides information on 75 fundamental frequency (f_o) extraction and "Glottal Closure Instant" (GCI) extimation methods, introduced in the last 50 years. Where possible, a link to available source code or software applications is provided. Given the sheer number of published material, the list does not claim to be comprehensive. We apologize to all authors whose algorithms we might have omitted.

ID	Name	Citation	fo/GCI	URL
N/A	a nonlinear algorithm for epoch marking in speech signals using poincare maps	Mann and McLaughlin (1998)	fo	
N/A	event-based instantaneous fundamental frequency estimation	B. Yegnanarayana and K. Murty (2009)	fo	
N/A	harmonics frequency estimation based on instantaneous frequency	Abe, Kobayashi, and Imai (1995)	fo	
N/A	hilbert envelope of linear prediction residual	Guruprasad, B. Yegnanarayana, and K. Sri Rama Murty (2007)	GCI	
N/A	hidden Markov-model multipitch tracking algorithm	Mingyang Wu, DeLiang Wang, and Brown (2003)	fo	
N/A	improved time domain pitch detection algorithm for pathological voice	Jamludin et al. (2012)	fo	
N/A	maximum likelihood harmonic matching and hidden Markov models	Doval and Rodet (1993)	fo	
N/A	method to determine the instants of significant excitation using the average group delay characteristics of minimum phase signals	Smits and B. Yegnanarayana (1995)	GCI	
N/A	multiband statistical learning	Sha, Burgoyne, and Saul (2004)	fo	
N/A	optimum comb method	Moorer (1974)	fo	
N/A	period histogram	Schroeder (1968)	fo	
N/A	Poincaré sections for pitch mark determination	Hagmüller and Kubin (2005)	fo	
N/A	real time harmonic pitch detector	Seneff (1978)	fo	

ID	Name	Citation	fo/GCI	URL
N/A	robust pitch determination using nonlinear state-space embedding	Terez (2002)	fo	
N/A	spectral autocorrelation method	Lahat, Niederjohn, and Krubsack (1987)	fo	
N/A	spectral equalization LPC method using Newton's transformation	Atal, unpublished, cited in L. R. Rabiner and Crochiere (1976)	fo	
N/A	statistical pitch detection algorithm	Y.-R. Wang, Wong, and Tsao (2002)	fo	
N/A	synthesis-based method for pitch extraction	Paliwal and P. Rao (1983)	fo	
N/A	tunable IIR filter	Lane (1990)	fo	
N/A	two-level autocorrelation method	Mitev and Hadjitodorov (2003)	fo	
(e)SRPD	Super-Resolution Pitch Determinator	Medan, Yair, and Chazan (1991)	fo	URL
ACF-AMDF	pitch detection Scheme based on ACF and AMDF	Kumar, Bhattacharya, and Patel (2014)	fo	
AGCD	Approximate Greatest Common Devisor algorithm	Sreenivas and P. V. S. Rao (1979)	fo	
AMDF	Average Magnitude Difference Function	Ross et al. (1974)	fo	
ASDF	Average Squared Difference Function	Nguyen and Imai (1977)	fo	
AUTOC	modified autocorrelation method using clipping	Dubnowski, Schafer, and L. Rabiner (1976)	fo	
BAC	Biased Auto-Correlation	Sondhi (1968)	fo	
CATE	Circular Autocorrelated Temporal Excitation	Di Martino and Laprie (1999)	fo	
CC-AMDF	Cross-Correlation AMDF	Chong Un and Shih-Chien Yang (1977)	fo	
CDP	cepstrum-based pitch detection algorithm	Luengo et al. (2007)	fo	
CEP	cepstrum	Noll (1967)	fo	URL
CWT	continuous wavelet transform	Manfredi et al. (2000)	fo	
DARD	data-reduction method	N. Miller (1975)	fo	
DLFT-PDA	discrete logarithmic Fourier transformation-pitch detection algorithm	Shapiro and C. Wang (2009)	fo	
DME–AR PSD	Dynamic Mean Evaluation Auto-Regressive Power Spectral Density method	Manfredi et al. (2000)	fo	
DYPSA	DYnamic programming Phase Slope Algorithm	Kounoudes, P. A. Naylor, and Brookes (2002)	GCI	
DECOM	DEgg Correlation-based method for Open quotient Measurement	Henrich et al. (2004)	GCI	URL
DyWT	event-based pitch detector using the dyadic wavelet transform	Kadambe and Boudreaux-Bartels (1992)	fo	

ID	Name	Citation	fo/GCI	URL
EFLPR	Epoch Filtering of LP Residual	Ananthapadmanabha and B. Yegnanarayana (1979)	GCI	
ESPS	Entropics Signal Processing System	unpublished. Code written by Shankar Narayan, Entropic Processing, Inc.; modified 1986 by David Burton	fo	URL
fxac (Speech Filing System)	auto-correlation	Mark Huckvale, unpublished	fo	URL
fxanal (Speech Filing System)	integrated pitch tracking algorithm	Secrest and Doddington (1983)	fo	URL
HE	Hilbert Envelope-based detection	Prasanna, Gupta, and B Yegnanarayana (2006)	GCI	
HIPEX	Harmonic Identification Pitch EXtraction	R. L. Miller (1970)	fo	
HPS	Harmonic Product Spectrum	Schroeder (1968)	fo	
HS	Harmonic Sieve	Duifhuis, Willems, and Sluyter (1982)	fo	
HTC	Harmonic-Temporal structured Clustering	Le Roux et al. (2007)	fo	
LoMA	Lines of Maximum Amplitude (across all the scales in the wavelet transform)	Tuan and D'Alessandro (1999)	GCI	
ML	Maximum Likelihood pitch estimator	Wise, Caprio, and Parks (1976)	fo	
NCCF	Normalized Cross Correlation	Atal and Saroop (1968)	fo	
NDF	Nearly Defect-Free f0 trajectory extraction	Kawahara, De Cheveigne, et al. (2005)	fo	URL
PEFAC	Pitch Estimation Filter with Amplitude Compression	Gonzalez and Brookes (2014)	fo	URL
PPROC	Parallel-PROCessing time-domain method	Gold and L Rabiner (1969)	fo	
Praat AC	Auto-Correlation	Boersma (1993)	fo	URL
Praat CC	forward Cross-Correlation analysis	N/A	fo	URL
RAPT	Robust Algorithm for Pitch Tracking	Talkin (1995)	fo	URL
REAPER	Robust Epoch And Pitch EstimatoR	David Talkin, Google Inc. - unpublished	GCI	URL
SAFE	Statistical Algorithm for F0 Estimation	Chu and Alwan (2012)	fo	
SAPD	Semi-Automatic Pitch Detector	McGonegal, L. Rabiner, and Rosenberg (1975)	fo	
SE-VQ	SEDREAMS algorithm (SE) modified to better handle voice qualities (VQ) resulting from different phonation types	Kane and Gobl (2013)	GCI	URL
SEDREAMS	Speech Event Detection using the Residual Excitation And a Mean-based Signal	Drugman, Drugman, and Alwan (2011)	GCI	URL

ID	Name	Citation	fo/GCI	URL
SHAPE	Smooth Harmonic Average Peak-to-valley Envelope	Camacho and Harris (2008)	fo	
SHR	Subharmonic to Harmonic Ratio	Sun (2000)	fo	URL
SHS	Sub-Harmonic Summation	Hermes (1988)	fo	URL
SIFT	Simple Inverse Filter Tracking	Markel (1972)	fo	
SIGMA	Singularity in EGG by Multiscale Analysis	M. Thomas and P. Naylor (2009)	GCI	URL
SPINET	Spatial PItch NETwork	Cohen, Grossberg, and Wyse (1995)	fo	
SRH	Summation of Residual Harmonics	Drugman, Drugman, and Alwan (2011)	fo	URL
SVD	Singular Value Decomposition	Ma, Kamp, and Willems (1994)	GCI	
SWIPE	Sawtooth Waveform Inspired Pitch Estimator	Camacho and Harris (2008)	fo	URL
TEMPO (STRAIGHT)	fixed point analysis of frequency to instantatneous frequency mapping	Kawahara, Kawahara, et al. (1999)	fo	URL
YAAPT	Yet Another Algorithm for Pitch Tracking	Kasi and Zahorian (2002)	fo	
YAGA	Yet Another GCI Algorithm	M. R. P. Thomas, Gudnason, and P. A. Naylor (2012)	GCI	
YIN	from "yin" and "yang" of oriental philosophy: interplay between autocorrelation and cancellation	De Cheveigné and Kawahara (2002)	fo	URL
ZFR	Zero Frequency Resonator-based method	K Sri Rama Murty and B Yegnanarayana (2008)	GCI	

Table 1: Non-exhaustive overview of algorithms for f_o and GCI detection.

References

Abe, T., T. Kobayashi, and S. Imai (1995). "Harmonics tracking and pitch extraction based on instantaneous frequency". In: *1995 International Conference on Acoustics, Speech, and Signal Processing*. Vol. 1. IEEE, pp. 756–759. ISBN: 0-7803-2431-5. DOI: 10.1109/ICASSP.1995.479804. URL: http://ieeexplore.ieee.org/document/479804/ (cit. on p. 1).

Ananthapadmanabha, T. and B. Yegnanarayana (1979). "Epoch extraction from linear prediction residual for identification of closed glottis interval". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 27.4, pp. 309–319. ISSN: 0096-3518. DOI: 10.1109/TASSP.1979.1163267. URL: http://ieeexplore.ieee.org/document/1163267/ (cit. on p. 3).

Atal and Bishnu Saroop (1968). *Automatic speaker recognition based on pitch contours*. URL: http://dl.acm.org/citation.cfm?id=904834 (cit. on p. 3).

Boersma, Paul (1993). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound". In: *Proceedings of the institute of phonetic sciences*. Vol. 17. 1193. Amsterdam, pp. 97–110 (cit. on p. 3).

Camacho, Arturo and John G. Harris (2008). "A sawtooth waveform inspired pitch estimator for speech and music". In: *The Journal of the Acoustical Society of America* 124.3, pp. 1638–1652. ISSN: 0001-4966. DOI: 10.1121/1.2951592. URL: http://www.ncbi.nlm.nih.gov/pubmed/19045655%20http://asa.scitation.org/doi/10.1121/1.2951592 (cit. on p. 4).

Chong Un and Shih-Chien Yang (1977). "A pitch extraction algorithm based on LPC inverse filtering and AMDF". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 25.6, pp. 565–572. ISSN: 0096-3518. DOI: 10.1109/TASSP.1977.1163005. URL: http://ieeexplore.ieee.org/document/1163005/ (cit. on p. 2).

Chu, Wei and Abeer Alwan (2012). "SAFE: A Statistical Approach to F0 Estimation Under Clean and Noisy Conditions". In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.3, pp. 933–944. ISSN: 1558-7916. DOI: 10.1109/TASL.2011.2168518. URL: http://ieeexplore.ieee.org/document/6021334/ (cit. on p. 3).

Cohen, M A, S Grossberg, and L L Wyse (1995). "A spectral network model of pitch perception." In: *The Journal of the Acoustical Society of America* 98.2 Pt 1, pp. 862–79. ISSN: 0001-4966. URL: http://www.ncbi.nlm.nih.gov/pubmed/7642825 (cit. on p. 4).

De Cheveigné, Alain and Hideki Kawahara (2002). "YIN, a fundamental frequency estimator for speech and music". In: *The Journal of the Acoustical Society of America* 111.4, pp. 1917–1930 (cit. on p. 4).

Di Martino, Joseph and Yves Laprie (1999). "An efficient F0 determination algorithm based on the implicit calculation of the autocorrelation of the temporal excitation signal". In: *6th European Conference on Speech Communication & Technology-EUROSPEECH'99*, 4–p (cit. on p. 2).

Doval, B. and X. Rodet (1993). "Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMMs". In: *IEEE International Conference on Acoustics Speech and Signal Processing*. IEEE, 221–224 vol.1. ISBN: 0-7803-0946-4. DOI: 10.1109/ICASSP.1993.319095. URL: http://ieeexplore.ieee.org/document/319095/ (cit. on p. 1).

Drugman, Thomas, Thomas Drugman, and Abeer Alwan (2011). "Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics". In: *Interspeech*. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.228.1866 (cit. on pp. 3, 4).

Dubnowski, J., R. Schafer, and L. Rabiner (1976). "Real-time digital hardware pitch detector". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 24.1, pp. 2–8. ISSN: 0096-3518. DOI: 10.1109/TASSP.1976.1162765. URL: http://ieeexplore.ieee.org/document/1162765/ (cit. on p. 2).

Duifhuis, Hendrikus, Lei F Willems, and RJ Sluyter (1982). "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception". In: *The Journal of the Acoustical Society of America* 71.6, pp. 1568–1580 (cit. on p. 3).

Gold, B and L Rabiner (1969). "Parallel processing techniques for estimating pitch periods of speech in the time domain." In: *The Journal of the Acoustical Society of America* 46.2, pp. 442–8. ISSN: 0001-4966. URL: http://www.ncbi.nlm.nih.gov/pubmed/5804116 (cit. on p. 3).

Gonzalez, Sira and Mike Brookes (2014). "PEFAC - A Pitch Estimation Algorithm Robust to High Levels of Noise". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22.2, pp. 518–530. ISSN: 2329-9290. DOI: 10.1109/TASLP.2013.2295918. URL: http://ieeexplore.ieee.org/document/6701334/ (cit. on p. 3).

Guruprasad, S., B. Yegnanarayana, and K. Sri Rama Murty (2007). "Detection of instants of glottal closure using characteristics of excitation source". In: *Interspeech*, pp. 554–557 (cit. on p. 1).

Hagmüller, Martin and Gernot Kubin (2005). "Poincaré Sections for Pitch Mark Determination". In: *ITRW on Non-Linear Speech Processing (NOLISP 05)*. Barcelona, Spain, pp. 1–7 (cit. on p. 1).

Henrich, Nathalie et al. (2004). "On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation". In: *The Journal of the Acoustical Society of America* 115.3, pp. 1321–1332 (cit. on p. 2).

Hermes, Dik J (1988). "Measurement of pitch by subharmonic summation". In: *The journal of the acoustical society of America* 83.1, pp. 257–264 (cit. on p. 4).

Jamludin, Mohd Redzuan et al. (2012). "An Improved Time Domain Pitch Detection Algorithm for Pathological Voice". In: *American Journal of Applied Sciences* 9.1, pp. 93–102. ISSN: 1546-9239. DOI: 10.3844/ajassp.2012.93.102 (cit. on p. 1).

Kadambe, S. and G.F. Boudreaux-Bartels (1992). "Application of the wavelet transform for pitch detection of speech signals". In: *IEEE Transactions on Information Theory* 38.2, pp. 917–924. ISSN: 0018-9448. DOI: 10.1109/18.119752. URL: http://ieeexplore.ieee.org/document/119752/ (cit. on p. 2).

Kane, John and Christer Gobl (2013). "Evaluation of glottal closure instant detection in a range of voice qualities". In: *Speech Communication* 55.2, pp. 295–314. ISSN: 0167-6393. DOI: 10.1016/j.specom.2012.08.011. URL: http://www.sciencedirect.com/science/article/pii/S0167639312001033 (cit. on p. 3).

Kasi, Kavita and Stephen A. Zahorian (2002). "Yet Another Algorithm for Pitch Tracking". In: *IEEE International Conference on Acoustics Speech and Signal Processing*. IEEE, pp. I–361–I–364. ISBN: 0-7803-7402-9. DOI: 10.1109/ICASSP.2002.5743729. URL: http://ieeexplore.ieee.org/document/5743729/ (cit. on p. 4).

Kawahara, Hideki, Alain De Cheveigne, et al. (2005). "Nearly Defect-free F0 Trajectory Extraction for Expressive Speech Modifications based on STRAIGHT". In: *Interspeech*, pp. 537–540 (cit. on p. 3).

Kawahara, Hideki, Hideki Kawahara, et al. (1999). "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of f0 and periodicity". In: *IN PROC. OF EUROSPEECH*, pp. 2781–2784. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.470.3349 (cit. on p. 4).

Kounoudes, Anastasis, Patrick A. Naylor, and Mike Brookes (2002). "The DYPSA algorithm for estimation of glottal closure instants in voiced speech". In: *ICASSP*, pp. 349–352 (cit. on p. 2).

Kumar, Sandeep, S. Bhattacharya, and Premanand Patel (2014). "A new pitch detection scheme based on ACF and AMDF". In: *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*. IEEE, pp. 1235–1240. ISBN: 978-1-4799-3914-5. DOI: 10.1109/ICACCCT.2014.7019296. URL: http://ieeexplore.ieee.org/document/7019296/ (cit. on p. 2).

Lahat, M., R. Niederjohn, and D. Krubsack (1987). "A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 35.6, pp. 741–750. ISSN: 0096-3518. DOI: 10.1109/TASSP.1987.1165224. URL: http://ieeexplore.ieee.org/document/1165224/ (cit. on p. 2).

Lane, John E. (1990). "Pitch Detection Using a Tunable IIR Filter". In: *Computer Music Journal* 14.3, p. 46. ISSN: 0148-9267. DOI: 10.2307/3679959. URL: http://www.jstor.org/stable/3679959?origin=crossref (cit. on p. 2).

Le Roux, Jonathan et al. (2007). "Single and Multiple F0 Contour Estimation Through Parametric Spectrogram Modeling of Speech in Noisy Environments". In: *IEEE Transactions on Audio, Speech and Language Processing* 15.4, pp. 1135–1145. ISSN: 1558-7916. DOI: 10.1109/TASL.2007.894510. URL: http://ieeexplore.ieee.org/document/4156204/ (cit. on p. 3).

Luengo, Iker et al. (2007). "Evaluation of Pitch Detection Algorithms Under Real Conditions". In: *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*. IEEE, pp. IV–1057–IV–1060. ISBN: 1-4244-0727-3. DOI: 10.1109/ICASSP.2007.367255. URL: http://ieeexplore.ieee.org/document/4218286/ (cit. on p. 2).

Ma, Changxue, Yves Kamp, and Lei F Willems (1994). "A Frobenius norm approach to glottal closure detection from the speech signal". In: *IEEE Transactions on Speech and Audio Processing* 2.2, pp. 258–265 (cit. on p. 4).

Manfredi, C et al. (2000). "A comparative analysis of fundamental frequency estimation methods with application to pathological voices." In: *Medical engineering & physics* 22.2, pp. 135–47. ISSN: 1350-4533. URL: http://www.ncbi.nlm.nih.gov/pubmed/10854967 (cit. on p. 2).

Mann, I and S McLaughlin (1998). "A nonlinear algorithm for epoch marking in speech signals using poincare maps". In: *9th European Signal Processing Conference (EUSIPCO 1998)*, pp. 1–4 (cit. on p. 1).

Markel, John (1972). "The SIFT algorithm for fundamental frequency estimation". In: *IEEE Transactions on Audio and Electroacoustics* 20.5, pp. 367–377 (cit. on p. 4).

McGonegal, C., L. Rabiner, and A. Rosenberg (1975). "A semiautomatic pitch detector (SAPD)". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 23.6, pp. 570–574. ISSN: 0096-3518. DOI: 10.1109/TASSP.1975.1162750. URL: http://ieeexplore.ieee.org/document/1162750/ (cit. on p. 3).

Medan, Y., E. Yair, and D. Chazan (1991). "Super resolution pitch determination of speech signals". In: *IEEE Transactions on Signal Processing* 39.1, pp. 40–48. ISSN: 1053-587X. DOI: 10.1109/78.80763. URL: http://ieeexplore.ieee.org/document/80763/ (cit. on p. 2).

Miller, N. (1975). “Pitch detection by data reduction”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 23.1, pp. 72–79. ISSN: 0096-3518. DOI: 10.1109/TASSP.1975.1162642. URL: http://ieeexplore.ieee.org/document/1162642/ (cit. on p. 2).
Miller, R L (1970). “Performance characteristics of an experimental harmonic identification pitch extraction (HIPEX) system.” In: *The Journal of the Acoustical Society of America* 47.6, pp. 1593–601. ISSN: 0001-4966. URL: http://www.ncbi.nlm.nih.gov/pubmed/5426310 (cit. on p. 3).
Mingyang Wu, DeLiang Wang, and G.J. Brown (2003). “A multipitch tracking algorithm for noisy speech”. In: *IEEE Transactions on Speech and Audio Processing* 11.3, pp. 229–241. ISSN: 1063-6676. DOI: 10.1109/TSA.2003.811539. URL: http://ieeexplore.ieee.org/document/1208292/ (cit. on p. 1).
Mitev, Petar and Stefan Hadjitodorov (2003). “Fundamental frequency estimation of voice of patients with laryngeal disorders”. In: *Information Sciences* 156.1, pp. 3–19. ISSN: 0020-0255. DOI: 10.1016/S0020-0255(03)00161-0. URL: http://www.sciencedirect.com/science/article/pii/S0020025503001610 (cit. on p. 2).
Moorer, J. (1974). “The optimum comb method of pitch period analysis of continuous digitized speech”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 22.5, pp. 330–338. ISSN: 0096-3518. DOI: 10.1109/TASSP.1974.1162596. URL: http://ieeexplore.ieee.org/document/1162596/ (cit. on p. 1).
Murty, K Sri Rama and B Yegnanarayana (2008). “Epoch extraction from speech signals”. In: *IEEE Transactions on Audio, Speech, and Language Processing* 16.8, pp. 1602–1613 (cit. on p. 4).
Nguyen, L.P. and S. Imai (1977). “Vocal pitch detection using generalized distance function associated with a voiced-unvoiced decision logic”. In: *Bull. P.M.E. (T.I.T.)* 39, pp. 11–21 (cit. on p. 2).
Noll, A Michael (1967). “Cepstrum pitch determination”. In: *The journal of the acoustical society of America* 41.2, pp. 293–309 (cit. on p. 2).
Paliwal, K.K. and P.V.S. Rao (1983). “A synthesis-based method for pitch extraction”. In: *Speech Communication* 2.1, pp. 37–45. ISSN: 0167-6393. DOI: 10.1016/0167-6393(83)90062-6. URL: http://linkinghub.elsevier.com/retrieve/pii/0167639383900626 (cit. on p. 2).
Prasanna, SR Mahadeva, Cheedella S Gupta, and B Yegnanarayana (2006). “Extraction of speaker-specific excitation information from linear prediction residual of speech”. In: *Speech Communication* 48.10, pp. 1243–1261 (cit. on p. 3).
Rabiner, L R and R E Crochiere (1976). “On the design of all-pass signals with peak amplitude”. In: *The Bell System Technical Journal* 55.4, pp. 395–407 (cit. on p. 2).
Ross, M et al. (1974). “Average magnitude difference function pitch extractor”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 22.5, pp. 353–362 (cit. on p. 2).
Schroeder, M R (1968). “Period histogram and product spectrum: new methods for fundamental-frequency measurement.” In: *The Journal of the Acoustical Society of America* 43.4, pp. 829–34. ISSN: 0001-4966. URL: http://www.ncbi.nlm.nih.gov/pubmed/5645832 (cit. on pp. 1, 3).
Secrest, Bruce and GEORGE Doddington (1983). “An integrated pitch tracking algorithm for speech systems”. In: *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'83.* Vol. 8. IEEE, pp. 1352–1355 (cit. on p. 3).
Seneff, S. (1978). “Real-time harmonic pitch detector”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.4, pp. 358–365. ISSN: 0096-3518. DOI: 10.1109/TASSP.1978.1163118. URL: http://ieeexplore.ieee.org/document/1163118/ (cit. on p. 1).
Sha, F., J.A. Burgoyne, and L.K. Saul (2004). “Multiband statistical learning for f/sub 0/ estimation in speech”. In: *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 5. IEEE, pp. V–661–4. ISBN: 0-7803-8484-9. DOI: 10.1109/ICASSP.2004.1327197. URL: http://ieeexplore.ieee.org/document/1327197/ (cit. on p. 1).
Shapiro, Ari Daniel and Chao Wang (2009). “A versatile pitch tracking algorithm: From human speech to killer whale vocalizations”. In: *The Journal of the Acoustical Society of America* 126.1, pp. 451–459. ISSN: 0001-4966. DOI: 10.1121/1.3132525. URL: http://www.ncbi.nlm.nih.gov/pubmed/19603902%20http://asa.scitation.org/doi/10.1121/1.3132525 (cit. on p. 2).
Smits, R. and B. Yegnanarayana (1995). “Determination of instants of significant excitation in speech using group delay function”. In: *IEEE Transactions on Speech and Audio Processing* 3.5, pp. 325–333. ISSN: 1063-6676. DOI: 10.1109/89.466662. URL: http://ieeexplore.ieee.org/document/466662/ (cit. on p. 1).
Sondhi, Mohan (1968). “New methods of pitch extraction”. In: *IEEE Transactions on audio and electroacoustics* 16.2, pp. 262–266 (cit. on p. 2).
Sreenivas, T. V. and P. V. S. Rao (1979). “Pitch extraction from corrupted harmonics of the power spectrum”. In: *The Journal of the Acoustical Society of America* 65.1, pp. 223–228. ISSN: 0001-4966. DOI: 10.1121/1.382239 (cit. on p. 2).
Sun, Xuejing (2000). “A pitch determination algorithm based on subharmonic-to-harmonic ratio”. In: *A A* 1000, p. 1 (cit. on p. 4).

Talkin (1995). “A robust algorithm for pitch tracking (RAPT)”. In: *Speech Coding and Synthesis*. Ed. by W. B. Kleijn and K. K. Paliwal. New York, NY: Elsevier Science Inc., pp. 495–518. URL: http://www.citeulike.org/user/gpk/article/2912571 (cit. on p. 3).

Terez, Dmitry E. (2002). “Robust pitch determination using nonlinear state-space embedding”. In: *IEEE International Conference on Acoustics Speech and Signal Processing*. IEEE, pp. I–345–I–348. ISBN: 0-7803-7402-9. DOI: 10.1109/ICASSP.2002.5743725. URL: http://ieeexplore.ieee.org/document/5743725/ (cit. on p. 2).

Thomas, Mark R. P., Jon Gudnason, and Patrick A. Naylor (2012). “Estimation of Glottal Closing and Opening Instants in Voiced Speech Using the YAGA Algorithm”. In: *IEEE Transactions on Audio, Speech, and Language Processing* 20.1, pp. 82–91. ISSN: 1558-7916. DOI: 10.1109/TASL.2011.2157684. URL: http://ieeexplore.ieee.org/document/5784321/ (cit. on p. 4).

Thomas, M.R.P. and P.A. Naylor (2009). “The SIGMA Algorithm: A Glottal Activity Detector for Electroglottographic Signals”. In: *IEEE Transactions on Audio, Speech, and Language Processing* 17.8, pp. 1557–1566. ISSN: 1558-7916. DOI: 10.1109/TASL.2009.2022430. URL: http://ieeexplore.ieee.org/document/4912310/ (cit. on p. 4).

Tuan, Vu Ngoc and Christophe D’Alessandro (1999). “Robust Glottal Closure Detection Using The Wavelet Transform.” In: *IN PROCEEDINGS OF THE EUROPEAN CONFERENCE ON SPEECH TECHNOLOGY*, pp. 2805–2808. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.3678 (cit. on p. 3).

Wang, Yih-Ru, I-Je Wong, and Teng-Chun Tsao (2002). “A statistical pitch detection algorithm”. In: *IEEE International Conference on Acoustics Speech and Signal Processing*. IEEE, pp. I–357–I–360. ISBN: 0-7803-7402-9. DOI: 10.1109/ICASSP.2002.5743728. URL: http://ieeexplore.ieee.org/document/5743728/ (cit. on p. 2).

Wise, J., J. Caprio, and T. Parks (1976). “Maximum likelihood pitch estimation”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 24.5, pp. 418–423. ISSN: 0096-3518. DOI: 10.1109/TASSP.1976.1162852. URL: http://ieeexplore.ieee.org/document/1162852/ (cit. on p. 3).

Yegnanarayana, B. and K.S.R. Murty (2009). “Event-Based Instantaneous Fundamental Frequency Estimation From Speech Signals”. In: *IEEE Transactions on Audio, Speech, and Language Processing* 17.4, pp. 614–624. ISSN: 1558-7916. DOI: 10.1109/TASL.2008.2012194. URL: http://ieeexplore.ieee.org/document/4802170/ (cit. on p. 1).